



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

What Can We Learn About Effective Mathematics Teaching?

Citation for published version:

Dieterle, S, Guarino, CM, Bargagliotti, AE & Mason, WM 2013, 'What Can We Learn About Effective Mathematics Teaching? A Framework for Estimating Causal Effects using Longitudinal Survey Data.', *Journal of Research on Educational Effectiveness*, vol. 6, no. 2, 2, pp. 164-198.
<https://doi.org/10.1080/19345747.2012.706695>

Digital Object Identifier (DOI):

[10.1080/19345747.2012.706695](https://doi.org/10.1080/19345747.2012.706695)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

Published In:

Journal of Research on Educational Effectiveness

Publisher Rights Statement:

© Dieterle, S., Guarino, C. M., Bargagliotti, A. E., & Mason, W. M. (2013). What Can We Learn About Effective Mathematics Teaching?: A Framework for Estimating Causal Effects using Longitudinal Survey Data. *Journal of Research on Educational Effectiveness*, 6(2), 164-198. [2]. 10.1080/19345747.2012.706695

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





California Center for Population Research
University of California - Los Angeles

What Can We Learn about Effective Early Mathematics Teaching? A Framework for Estimating Causal Effects using Longitudinal Survey Data

Cassandra M. Guarino
Steven G. Dieterle
Anna E. Bargagliotti
William M. Mason

PWP-CCPR-2011-013

***California Center for Population Research
On-Line Working Paper Series***

What Can We Learn about Effective Early Mathematics Teaching? A Framework for Estimating Causal Effects using Longitudinal Survey Data

By Cassandra M. Guarino,¹ Steven G. Dieterle,² Anna E. Bargagliotti,³ and William M. Mason⁴

¹Indiana University

²Michigan State University

³Loyola Marymount University

⁴UCLA

Acknowledgment: This work was supported in part by the National Institute of Child Health and Human Development (NICHD) grant 5R03HD57510-2, “The Impact of Teacher Qualifications and Instructional Practices on Reading and Mathematics Achievement in Young Children,” C. M. Guarino, Principal Investigator. Steven Dieterle’s participation was funded in part by Pre-Doctoral Training Grant R305B090011 from the Institute of Education Sciences (IES), U.S. Department of Education. The views expressed here are those of the authors and do not necessarily reflect the views of NICHD or IES.

Version 12/31/11

Abstract

This study investigates the impact of teacher characteristics and instructional strategies on the mathematics achievement of students in kindergarten and first grade and tackles the question of how best to use longitudinal survey data to elicit causal inference in the face of potential threats to validity due to nonrandom assignment to treatment. We develop a step-by-step approach to selecting a modeling and estimation strategy and find that teacher certification and courses in methods of teaching mathematics have a slightly negative effect on student achievement in kindergarten, whereas postgraduate education has a positive effect in first grade. Various teaching modalities, such as working with counting manipulatives, using math worksheets, and completing problems on the chalkboard, have positive effects on achievement in kindergarten, and pedagogical practices relating to explaining problem-solving and working on problems from textbooks have positive effects on achievement in first grade. We show that the conclusions drawn depend on the estimation and modeling choices made and that several prior studies of teacher effects using longitudinal survey data likely neglected important features needed to establish causal inference.

1. Introduction

This study investigates the impact of teacher characteristics and instructional strategies on the mathematics achievement of students in kindergarten and first grade. Understanding the factors that make some teachers more effective than others is vital to achieving and supporting high quality instruction. Early teaching, in particular, can be crucial to the future academic progress of children, as well as in determining later economic well being and other nonacademic outcomes (Barnett, 1995; Currie & Thomas, 2000; Kilpatrick, Swafford & Findell, 2001; Chetty et al., 2010).

Evidence of “what works” in elementary mathematics instruction can be obtained from multiple sources: experiments, observations, administrative data, and surveys. This study utilizes survey data and provides a framework to guide the estimation of causal effects in nonexperimental settings. Although experimental evidence of effects is generally considered the gold standard, true educational experiments are rare, centered primarily around interventions, and difficult to impose or implement. In addition, they are generally conducted on a small scale, yielding potentially ambiguous conclusions regarding the effects of scaled-up interventions. Classroom observations provide an in-depth picture of teaching but are costly to conduct on a large scale and can be difficult to parse into identifiable, quantifiable elements of instruction that overcome inter-rater reliability issues. Administrative data are often used to uncover associations between factors like teacher training and student achievement. However, these data contain only the most basic teacher and student characteristics required for reporting and compliance purposes and often suffer from missing, and sometimes inaccurate, data records. Large sample survey data, if representative of the population of interest and sufficiently detailed, can represent an improvement over administrative data. They are less costly to collect than classroom

observations on a large scale, and they can provide information on treatments that cannot be investigated through experiments. Much of the body of knowledge on teacher effectiveness consists of estimates derived from survey data, such as National Educational Longitudinal Study (NELS) and Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). However, these data have significant limitations, as well. They generally rely on self-reported activities and characteristics and often suffer from item non-response, even when overall survey response rates are high. In addition, sample attrition in longitudinal surveys can pose a threat to representativeness if it is nonrandom.

The fundamental disadvantage of survey data with respect to experimental data, however, is that students are not randomly assigned to treatments, rendering causal inference inherently difficult—a disadvantage shared with observational and administrative data. Survey data can compensate for this handicap to varying degrees, however, by providing a rich set of control variables that might reduce omitted variable bias in estimates. This paper tackles the question of how best to use longitudinal survey data to elicit causal inference with respect to teacher-related factors impacting early mathematics achievement in the face of potential threats to validity due to nonrandom assignment to treatment. Researchers have several tools at their disposition to deal with these threats. Methodological decisions can be complex, however, and, as we show, results can be sensitive to the approach selected. As such, it is imperative to understand why inconsistencies occur and which methods are best in a given situation.

The goals of this paper are thus twofold: (1) to lay out a careful approach for selecting an appropriate model and estimation method to investigate teacher effects using longitudinal survey data and (2) to apply this approach in answering our specific research question—i.e., the extent to which the observable background characteristics and instructional practices of kindergarten

and first grade teachers produce gains in the mathematics achievement of their students. We use data from ECLS-K, a nationally representative sample of kindergarteners followed over time. The data include student assessments in mathematics and reading at each wave as well as detailed information from parents, teachers, and school administrators and are therefore well suited to an investigation of our research question.

Through a step-by-step analysis of the data, we select a modeling and estimation strategy. Our findings indicate that teacher certification and courses in methods of teaching mathematics have a slightly negative effect on student achievement in kindergarten, whereas postgraduate education has a positive effect in first grade. Various teaching modalities, such as working with counting manipulatives, using math worksheets, and completing problems on the chalkboard, have positive effects on achievement in kindergarten, and pedagogical practices relating to explaining problem-solving and working on problems from textbooks have positive effects on achievement in first grade.

We find that the models and estimators previously employed to estimate teacher characteristic and practice effects using longitudinal survey data likely neglected important features needed to establish causal inference. Importantly, we show that the conclusions drawn depend on the estimation and modeling choices made, underscoring the importance of setting out a clear strategy for choosing among the many possibilities available.

This paper is organized as follows. In section 2, we outline a framework for selecting a model and method for estimating teacher effects using longitudinal survey data. Then, in section 3, we review the relevant literature pertaining to our specific research question—i.e., what teacher characteristics and practices affect student achievement in the early grades?—with

particular attention to the estimation methods used. Section 4 describes our data, section 5 outlines our methods, section 6 presents and discusses results, and section 7 concludes.

2. Modeling and Estimation Framework

Models

A very general cumulative effects model views current achievement as a function of all relevant current and past inputs, a student specific effect, and a random error term (Hanushek, 1979; Todd & Wolpin, 2003; Harris, Sass, & Semykina, 2010; Guarino, Reckase, & Wooldridge, 2011):

$$(1) \quad Y_{ijst} = f_t(X_{it}, \dots, X_{i0}, Z_{it}, \dots, Z_{i0}, c_i, u_{ijst})$$

where

Y_{ijst} = achievement of child i with teacher j in school s at time t
 f_t = time-varying education production function relating inputs to achievement
 X_{it} = time-varying child, family, and neighborhood inputs for child i in period t
 Z_{it} = time-varying schooling inputs (such as teaching practices, peer effects, etc.)
 c_i = time-invariant child effect
 u_{ijst} = unobserved error term

Researchers make several assumptions to render (1) tractable for analysis. As a first step, the function f_t is generally assumed to be linear in the parameters. These assumptions yield the *linear cumulative effects* (LCE) model:

$$(2) \quad Y_{it} = \alpha_t + X_{it}\theta_0 + X_{it-1}\theta_1 + \dots + X_{i0}\theta_t + Z_{it}\beta_0 + Z_{it-1}\beta_1 + \dots + Z_{i0}\beta_t + \eta_t c_i + u_{it}$$

Our interest lies primarily in estimating the β parameters, which in this case convey the partial effects of teacher characteristics and practices on math achievement. Typically, data on all inputs prior to kindergarten—e.g., preschool or daycare characteristics—are unavailable, and the student effect c_i , sometimes thought of as unmeasured innate student ability or motivation, is

generally unobservable. To address the lack of prior inputs, it is customary to impose an additional assumption: namely, that the effect of the inputs decays at a geometric rate equal to λ . In terms of the parameters, this assumption requires $\beta_s = \lambda^s \beta_0$ and $\theta_s = \lambda^s \theta_0$ for $s = 1, \dots, T$. The geometric decay assumption on the input effects allows one to eliminate the lagged inputs and rewrite equation (2) as a geometric distributed lag (GDL) model:¹

$$(3) \quad \begin{aligned} Y_{it} &= \alpha_i + \lambda Y_{it-1} + Z_{it} \beta_0 + X_{it} \theta_0 + \pi_i c_i + e_{it} \\ e_{it} &= u_{it} - \lambda u_{it-1} \end{aligned}$$

A second common restriction is that $\pi_i = \pi$, indicating that the child-specific effect has the same effect on achievement in every period. Note that under this assumption there is no loss of generality by denoting π as 1 in these models if a constant is present in the model.

$$(4) \quad Y_{it} = \alpha_i + \lambda Y_{it-1} + Z_{it} \beta_0 + X_{it} \theta_0 + c_i + e_{it}$$

Equation (4) is generally referred to as a *dynamic* linear model due to the presence of lagged achievement on the right-hand side. Estimation of (4) is generally feasible, as it requires only contemporaneous inputs and a lag of achievement. However, many researchers proceed to subtract prior achievement from both sides of the equation and estimate what is often referred to as a gain score model. This results in the following:

$$(5) \quad Y_{it} - Y_{it-1} = \alpha_i + Z_{it} \beta_0 + X_{it} \theta_0 + c_i + e_{it} - (1 - \lambda) Y_{it-1}$$

Of course, if $\lambda = 1$, then the piece of prior achievement left in the error term with a negative coefficient disappears. If not, it causes an omitted variables problem as well as negative serial correlation. Thus, in choosing this model specification, researchers are essentially

¹ Algebraically, (3) is derived from (2) by adding and subtracting λY_{it-1} from both sides of (2), imposing the GDL assumption, substituting and simplifying terms. See Guarino, Reckase, and Wooldridge (2011) and Harris, Sass, and Semykina (2010) for a detailed explanation of the geometric distributed lag assumption and its application to value-added modeling.

assuming that $\lambda = 1$ (i.e., that there is no decay in the impact of prior inputs on current achievement) or that the consequences of violating this assumption are unimportant in estimating the parameters of interest.

To relate equations (4) and the constrained version (5) directly to our research questions, we now rewrite them under all the assumptions hitherto mentioned using terms specific to our study:

$$(6) \quad Y_{ijst} = \lambda Y_{i,t-1} + \beta_1 PDG_{ijst} + \beta_2 TC_{ijst} + \beta_3 Class_{ijst} + \beta_4 X_{ijst} + \alpha_t + \delta_s + \tau_j + c_i + \varepsilon_{ijst}$$

$$(7) \quad Y_{ijst} - Y_{i,t-1} = \beta_1 PDG_{jst} + \beta_2 TC_{jst} + \beta_3 Class_{jst} + \beta_4 X_{ijst} + \alpha_t + \delta_s + \tau_j + c_i + \varepsilon_{ijst} + (\lambda - 1)Y_{i,t-1}$$

where

PDG = the set of pedagogical practices used by teacher j

TC = the set of background characteristics of teacher j

$CLASS$ = a set of classroom characteristics

X = a set of child- and family-related control variables

τ_j = an unobserved teacher-specific effect

δ_s = an unobserved school-specific effect

ε_{ijst} = a random error term

In our study, we use data from three time periods. ECLS-K assessed kindergarten and first grade children's achievement in the full sample in the fall of the kindergarten year ($t = 0$), in the spring of the kindergarten year ($t = 1$), and in the spring of the first grade year ($t = 2$).² In our modeling, the fall test score in kindergarten is lagged relative to the spring test score. For first grade, the spring kindergarten test score is lagged relative to the spring first grade test score.³ The characteristics of kindergarten teachers were recorded in the fall, and corrected in the spring if children's teachers changed. Information on teaching practices was recorded in the spring. For first grade teachers, all information was recorded in the spring.

² A subsample of first grade children was assessed in the fall; we do not use that partial wave.

³ Thus the intervals between current and lagged tests differ across grades. We take this into account by controlling for time elapsed between tests and, in some cases, by interacting lagged achievement with a grade indicator.

In these new equations (6) and (7), the composite error term ($\delta_s + \tau_j + c_i + \varepsilon_{ijst}$) contains unobserved school and teacher effects as well as the child effect c_i and the idiosyncratic term ε_{ijst} . By including the terms δ_s and τ_j , we allow for the possibility that in our focus on the specific teacher characteristics and pedagogical behaviors contained in *PDG* and *TC*, certain teacher and school-level factors that are relevant to predicting achievement may be omitted.⁴ It is important to note, however, that δ_s can be explicitly estimated through the inclusion of school dummy variables rather than left as a component of the unobserved error term, whereas fixed teacher effects τ_j —e.g., an unobserved teacher quality component—cannot be estimated separately from the observed teacher variables of interest unless we observe more than one class of students for each teacher.⁵

For notational simplicity, we do not include interactions in this model, although in our investigations, we explore the possibility that the effects of teacher characteristics and instructional practices on achievement differ across grades.

Estimators

Several estimation strategies can be applied to models (6) and (7). The estimators associated with these two models—the lag score and the gain score—are outlined below and summarized in Figure 1. Both models can be estimated using either cross-sections of data (i.e., data for each grade separately) or data that are pooled across kindergarten and first grade.

If the equations are estimated using cross-sections of data, two primary approaches are possible: ordinary least squares (OLS) estimation and maximum likelihood estimation with random effects assumptions. The OLS estimator will be consistent if student heterogeneity and the other error components are uncorrelated with either the input or output variables. We can

⁴ In our notation, we impose the simplifying assumptions that the unobserved school and teacher effects are constant across students and over time.

⁵ In ECLS-K, we have only one classroom per teacher. The possible effects of this limitation are later discussed.

relax that assumption with respect to δ_s , if need be, by putting in school dummy variables. In Figure 1, this set of choices is represented by the OLS portion of the tree on the left side of the diagram for cross-sectional data.

Efficiency gains may be possible using maximum likelihood if we exploit the nested structure of the data (in our case, children within classrooms within schools) and assume that the teacher and school terms (δ_s and τ_s) are normally distributed random effects. Such estimation strategies are often referred to as mixed or hierarchical linear models (HLM). This strategy effectively accounts for correlation among test scores for students with the same teacher and within the same school in the estimation of the pedagogy and teacher effects. Here again, we can treat the school effects as fixed rather than random by including school dummy variables, and use an HLM estimator with just teacher random effects (see the HLM portion of the tree diagram in Figure 1).

If the data are pooled across the grades, OLS can still be used on either model (6) or (7) under the same independence assumptions as in the cross-section data case (see the right portion of the figure under “panel”). However, panel data allow us to make use of approaches that deal with the presence of unobserved student heterogeneity.

In the lag score model (6), pooling across grades allows for the elimination of heterogeneity by first differencing and then instrumenting for the endogenous lagged test score gain with the twice-lagged test score. A common estimator that accomplishes this is the generalized method of moments (GMM) approach described in Arellano and Bond (1991), which we will henceforth refer to as AB. This approach not only accounts for unobserved student heterogeneity but also allows us the flexibility of leaving λ unconstrained. However, it relies on the assumptions that the error term in equation (6) is serially uncorrelated and that all other

inputs satisfy a “strict exogeneity” assumption—namely, that errors ε_{it} in one time period are uncorrelated with inputs in all other time periods (Wooldridge, 2002, ch. 10, pp.252-254).

In the gain score model (7), in which λ is constrained to equal 1, random effects (RE) or fixed effects (FE) estimators can be used to mitigate problems associated with student heterogeneity. RE assumes the child-specific heterogeneity and the inputs are uncorrelated, strict exogeneity, as well as a particular structure to the error covariance matrix, and may result in efficiency gains over OLS if these assumptions are met (Wooldridge 2002, ch. 10, pp. 252-254). FE estimation relaxes the assumption of zero correlation between heterogeneity and observed right-hand-side variables but maintains strict exogeneity. Note that the strict exogeneity assumption required by RE and FE is violated in model (6) because the lagged test score variable on the right-hand side is a function of the error term in the previous period. Thus, RE and FE estimators will be inconsistent for (6) and will only apply to the gain score model.

The modeling and estimation choices that we have described—18 in total and summarized in Figure 1—will be those considered in our study. In section 4, we outline a strategy for selecting the appropriate method. But first, we review prior survey-based research pertaining to our research question with careful attention to the methods chosen in these studies. Divergence of our findings from earlier findings may be traceable, at least in part, to the different choices of methods.

3. Prior survey-based research on the impact of teacher characteristics and teaching practices on student achievement in the early grades

Although a number of prior studies have used survey data to examine the relationship between student achievement and observable teacher characteristics and practices, using a variety of model specifications and estimation methods (see Wayne and Youngs, 2003, for a

review), relatively few have examined these relationships in the context of early elementary mathematics teaching. Rowan, Correnti, and Miller (2002) estimated a gain score model using HLM with both teacher and school random effects to analyze a longitudinal data set on elementary school children across the US in the early 1990s. They found no effect of teacher certification status or subject-matter preparation on achievement growth in mathematics but found a positive relationship between teaching experience and growth for students going from third to sixth grade and a negative relationship between advanced degrees and growth for students in both the early and upper elementary grades. Among a small set of pedagogical practices examined, only time spent on whole class instruction was positively related to mathematics achievement. Other studies focused on links between teacher-reported instructional practices and early mathematics achievement have tended to explore the use of “reform-based” or “standards-based” practices that emphasize problem solving and inquiry. Le et al. (2006), in a study of third grade students in five districts across the U.S. followed longitudinally for three years estimated a lag score model for achievement using OLS. They found weak and inconsistent relationships between student-centered practices and mathematics achievement. Reform-based practices showed positive effects on measures related to problem solving but negative effects on measures designed to capture a student’s grasp on mathematical procedures. Cohen & Hill (2000) also estimated a lag score model using OLS to study links between teacher responses on a survey administered to elementary teachers in California in 1994 and student achievement on the California Learning Assessment System during the same year. They found that teacher-reported frequency of use of reform-based practices was positively related to mathematics test scores among fourth-graders. Hamilton et al. (2003) conducted a meta-analytic synthesis of teacher survey data from grades three through seven at several National Science Foundation-funded

Systemic Reform Initiative sites to investigate the effect of reform-based teaching practices on mathematics and science learning. Their results were based on OLS regressions, although it is not clear whether gain or lag scores were used and specifications differed across sites depending on the availability of data. They combined several specific practice items into a “reform-based” scale and found small and weak but fairly consistent positive relationships between teachers’ use of reform-based practices and student achievement.

In addition to the studies cited above, three prior studies estimated the effects of teacher characteristics and instructional practices on the mathematics achievement of early elementary students using ECLS-K. All three relied on a single cross-section of the ECLS-K for the main analysis, with Guarino, Hamilton, Lockwood, and Rathbun (2006) and Bodovski and Farkas (2007) focusing on kindergarten and Parlady and Rumberger (2008) focusing on first grade. All employ HLM estimation.

Guarino et al. (2006), using a gain score model, found no evidence of a direct relationship between the background characteristics of teachers and student achievement but found that spending more time on subject was associated with relatively large gains in achievement. They constructed instructional practice scales, combining several practice measures into aggregate indexes using factor analysis. Among the scales designed to capture pedagogical approaches, those describing an emphasis on traditional practices and computation, measurement and advanced topics, advanced numbers and operations, and student-centered instruction (e.g., having students explain how problems were solved) were positively associated with mathematics achievement gains.

Bodovski and Farkas (2007) also focused on kindergarten and created instructional practice scales but utilized a lag score model in their analysis. They found that both traditional

and interactive approaches, along with practices that focused on advanced counting, practical math and single-digit operations were related to larger gains in achievement. On the other hand, spending additional time on basic numbers and shapes was found to be associated with lower achievement gains. They also mention obtaining similar results from an unreported fixed effects estimation which relies on the pooled kindergarten and first grade data available in the ECLS-K as evidence that their results approximate causal effects.

Parlady and Rumberger (2008) used a lag score model and focused on first grade rather than kindergarten. They found that the use of math worksheets and calendars raised student mathematics achievement, whereas the use of geometric manipulatives lowered it. They restricted their sample to a 30 percent subsample of students who were tested in the fall of first grade and used the fall exam score as an explanatory variable in the analysis.

The relationships found in these survey-based studies cannot be interpreted as causal unless the assumptions underlying the models and estimators used are met. In this paper, we analyze the ECLS-K data using a step-by-step approach to justify our modeling and estimation choices, the goal being to provide estimates with the best claim to causal inference. We then compare our results with those obtained in earlier studies using other methods.

4. Data

The ECLS-K selected a nationally representative sample of approximately 22,000 children who were enrolled in approximately 1,000 kindergarten programs in the United States, during the 1998-99 school year. The children were selected from both public and private kindergartens offering full-day and part-day programs. The sample consisted of children from different racial-ethnic and socioeconomic backgrounds and included an oversample of Asian children and private school kindergartners. The sample design for the ECLS-K was a dual-frame,

multistage sample. First, 100 Primary Sampling Units were selected (PSUs were counties or groups of counties). Schools within the PSUs were then selected; public schools from a public school frame and private schools from a private school frame. In the fall of 1998, approximately 23 kindergartners were selected within each of the sampled schools (Tourangeau et al., 2001).

The ECLS-K followed these children at various intervals through eighth grade. Three of the seven available waves of the data are utilized in this study. The first two waves of data were collected in the fall and spring of the 1998-1999 kindergarten year, respectively. The third and fourth waves were collected in the fall and spring of the first grade year, but the third wave (fall of the first grade year) was collected only on a relatively small (30 percent) sub-sample of the children and is therefore not used in this study. For this study, we restrict the data to the fall and spring kindergarten and spring first grade waves. The fifth, sixth, and seventh waves are excluded because they occur after intervals of two or three years—thus the variation in children’s learning gains is likely to be only loosely connected to the practices and abilities of contemporaneous teachers. In the three waves selected for the study, we make use of several categories of data—achievement assessments, teacher interviews, and student and family characteristics.

Achievement Assessments

Assessments that included cognitive components were conducted with the sampled children through one-on-one tests administered by trained individuals at each wave. The full achievement assessment used a computer-assisted personal interview and took approximately 50-70 minutes to complete. It included tests of reading and mathematics as well as other components that differed by wave (e.g., general knowledge in the kindergarten wave and science in the third grade). The test was untimed, and the kindergarten test required children to respond

verbally or through pointing; no writing was required. Each test was conducted using a two-stage design. The first stage consisted of a routing section that was administered to all students, and the second stage consisted of one of several alternative forms, the choice of which depended on the child's performance on the first stage. Only the assessments in mathematics are utilized in this study. The mathematics assessments had low, middle, and high difficulty second-stage options. The purpose of the adaptive design was to maximize accuracy of measurement and minimize administration time.⁶

The content of the mathematics assessments was selected to represent cognitive skills that are typically taught at each stage of development and that are important for the development of later proficiency (Rock & Pollack, 2002). Efforts were made to accommodate children who spoke a language other than English in the kindergarten and first grade assessments. Prior to administering these assessments, a language-screening test—the Oral Language Development Scale (OLDS)—was administered to those children identified from their school records (or by their teacher, if no school records were available) as coming from a home in which the primary language spoken was not English. Children whose performance exceeded an established cut score on the OLDS received the full English direct assessment in mathematics. Students who did not pass the OLDS but who spoke Spanish were given a translated form of the mathematics assessment. Various methods were used to confirm that the psychometric properties of the Spanish mathematics assessment were comparable to those for the English version (Rock & Pollack, 2002).

⁶ See the User's Manual for the ECLS-K Base Year Public-Use Data Files and Electronic Codebook, NCES 2001-029 (U.S. Department of Education, National Center for Education Statistics, 2001) or the ECLS-K Psychometric Report for Kindergarten through First Grade (Rock and Pollack, 2002) for a more complete description of the assessment procedures.

Three types of scores were reported for each test: (1) the number of questions answered correctly on the first-stage routing test, (2) item response theory (IRT) scale scores, and (3) standardized (t-scale) scores. The most appropriate of these for the purpose of this study are the IRT scores, because IRT scores are designed to make it possible to calculate scores that can be compared regardless of which second-stage form a child took in the adaptive test. They compensate for the possibility of a low-ability student guessing several items correctly. In addition, they make possible longitudinal measure of gain in achievement over time, even though the tests administered are not identical at each point (Tourangeau et al., 2001).

Teacher-level variables

Information on the teachers in both kindergarten and first grade was gathered in a set of self-administered paper-and-pencil questionnaires that included questions about their backgrounds and instructional practices. Background characteristics used in this study consisted of indicators for race/ethnicity, teaching experience, certification, educational attainment, and completion of courses in methods of teaching mathematics.⁷ Other relevant variables consisted of time spent on preparation, and, most importantly, a set of instructional practices described in the next section.

Instructional Practices

The spring teacher questionnaires include sets of items that address instructional practices in mathematics. The items address a wide range of practices that may occur in classrooms in the early grades and were selected to align with the skills tapped by the ECLS-K achievement assessments.

⁷ Based on exploratory analysis, (i) we trichotomized years of teaching experience as 0-4, 4-10, 10+; (ii) we dichotomized coursework as 0-2 versus 3+ courses.

Both the kindergarten and first grade teachers were asked very similar questions regarding their instructional practices; thus we were able to construct nearly identical sets of practices that apply to the two time periods. Specific pedagogical practices are listed as items in the ECLS-K teacher questionnaire under the question: “How often do children in the class do each of the following math activities?” The kindergarten teacher questionnaire includes 17 activities representing different pedagogical modalities. The first grade teacher questionnaire included the same items, with very few differences.⁸ We code teacher responses on all of these items to reflect days per month.⁹

In addition to these items, we include a measure of time spent on mathematics in our analyses. Teachers were asked how often they teach mathematics and how much time they spend on the subject on the days they teach it. We combined the responses to both questions to estimate the total hours per week a teacher reported spending on mathematics. In addition, teachers were asked the extent to which they utilized divided achievement grouping, without special reference to mathematics. This was coded as hours per week the student’s spent in such groups.

Content Coverage

In addition to the pedagogical variables, the surveys contain several items relating to content coverage. The stem question is: “How often is each of the following math skills taught in the class?” and 29 skill or content areas are then listed. We use these items as control variables to enable us to isolate the effect of pedagogical techniques holding constant content emphases that might align to a greater or lesser degree with the tests. They are recoded in a manner similar

⁸ The first grade questionnaire adds “Work on problems for which there are several appropriate methods or solutions” and “Do worksheet or workbook page emphasizing routine practice or drill” to the 17 kindergarten pedagogy questions.

⁹ We code the response categories for mathematics activities using what is essentially interval midpoint scaling: “never” → 0 days per month; “once a month or less” → 1 day per month; “two or three times a month” → 2.5 days per month; “once or twice a week” → 6 days per month; “three or four times a week” → 14 days a month; “daily” → 20 days per month. The metric assumes a standard of four weeks in a month and five working days per week.

to that of the pedagogy variables.¹⁰ Appendix Table 1 displays descriptive statistics for the teacher variables included in the model.

Classroom Characteristics

Teachers were also asked to describe demographic characteristics of their classes. Reported are class size and the percentages of children of different racial-ethnic groups and with disabilities. We include these in our analyses to control for differential teacher responses to variation in classroom composition.

Child and Family Variables

Student-specific variables used in our analyses include the number of days elapsed between tests, indicators for disability status, attending a full day kindergarten, whether the child is repeating kindergarten, and whether the child takes the test in Spanish. In addition, we include several controls that capture socioeconomic status (e.g., parent education and income), family behaviors (e.g., parental involvement, the number of extracurricular activities in which the child engages, how often the child reads, and the number of educational activities in which the child participates in the home), and family structure (e.g., single parent, number of siblings). ECLS-K provides a rich set of such variables, allowing us to capture child effects that are often omitted in administrative data. The full set of these controls and their descriptive statistics are also included in Appendix Table 1.

Sample Adjustments

Although item non-response for the variables in our study is typically low (e.g., around one to two percent for the pedagogy practices, as shown in Appendix Table 2), the combined

¹⁰ The response categories for the skills (content) items are the same as those for the activities (pedagogical modalities) items, except that the “never” category was named “not taught” and expanded into “taught at a higher grade level” and “children should already know.” We code the “not taught” categories as 0 times per month.

effect of missing item responses across all variables leads to a sample decrease of more than 60 percent in kindergarten and more than 65 percent in first grade. This drop in the number of observations hinders our ability to estimate the pedagogy and teacher characteristic effects with precision. To counter the loss of information stemming from item non-response, we used Royston's (2004) Stata implementation of chained multiple imputation (Van Buuren, Boshuizen & Knook, 1999) to impute missing values for all variables except student test scores. Forty imputed data sets were produced for each of three types of data—pooled data from both kindergarten and first grade, data from kindergarten only, and data from first grade only. The 40 pooled data sets were each composed of 21,232 student-year observations representing students with test score at all three waves. The separate kindergarten and first grade cross-sectional imputed data sets were composed of 16,356 and 11,780 student observations, respectively, including students with non-missing current and lagged test scores only in the particular grade imputed. Post-imputation estimation was carried out using Stata routines influenced by Carlin, Galati & Royston (2008). Following ECLS-K guidelines, we used sampling weights supplied with the data in the imputation to better approximate the initial population.

5. Methods

Here we outline a decision-making process to choose among the 18 model/estimation choices described in Section 2 and illustrated in Figure 1. To decide among these alternatives, we undertake a multi-step investigation. Each step pertains to one of the decisions needed: lag score versus gain score, cross-section versus panel, and ultimately choosing an estimator.

Gain Score vs. Lag Score

Deciding between a gain score or lag score model amounts to testing whether the assumption $\lambda=1$ in equation (6) is justifiable. In addition to merely observing the coefficient of

the lagged test score when estimating equation (6), we make use of a test proposed by Harris, Sass, and Semykina (2011). Applied to our problem, this test amounts to testing the joint significance of including a set of variables representing the first lags of all the inputs in the gain score equation (7) for first grade. Formally, this is a test of the null hypothesis that $\lambda=1$. The test is motivated by the idea that the lagged score, with the coefficient constrained to equal one, effectively serves as a sufficient statistic for past inputs. Therefore, if the gain score approach properly controls for past inputs, the included lagged inputs should not be statistically significant. It should be noted, however, that since we observe only two grades, the test of $\lambda=1$ will pertain only to first grade because there are no lags of inputs available for kindergarten.

Cross Section vs. Panel

The decision to use panel data versus separate cross-sections in our analyses is based on three considerations: whether the impact of teacher characteristics and practices changes across grades, whether there are precision gains due to increased sample sizes, and whether there is a need for panel data methods to eliminate time-invariant unobserved child heterogeneity, which may bias estimates of teacher effects related to *PDG* and *TC*.

To investigate whether the impact of teaching practices varies across grades, we estimate equation (6) or (7) (depending on whether we choose a gain score or a lag score model) using the pooled data and including interactions between a grade dummy and all teacher characteristics and practices. If several characteristics and practices interact significantly with grade, it would indicate that either the interactions should be included in any panel analyses or that cross-sectional regressions should be run separately by grade.

Possible precision gains from using the panel, due to the increased number of observations, might be a reason to prefer the panel regressions, but such gains are likely to be

quite small. As mentioned in the previous section, the grade-specific sample sizes in ECLS-K (N=16,356 for kindergarten and N=11,780 for grade 1) are likely large enough to mitigate concerns over lack of precision.

The most important driver of the decision regarding the panel versus the cross-sections is the question of whether unobserved child-specific heterogeneity c_i is likely to bias the teacher effect estimates. Only panel data methods (i.e., FE in the gain score model and AB in the lag score model) address this problem by eliminating the effects of heterogeneity. Assessing the influence of heterogeneity is not straightforward and the tools and tests at our disposal provide only suggestive evidence of the extent of the problem. Nevertheless, we can make use of the following set of procedures.

A common approach to assessing the influence of unobserved heterogeneity on the effect estimates is to use a Hausman (1978) test in which the coefficients from random and fixed effects estimators are compared, or a related variable addition test suggested by Wooldridge (Wooldridge, 2002, ch.10, p.288-291).¹¹ It should be noted, however, that in our achievement regression, the RE and FE estimators are consistent only for the gain score model; thus we can seek evidence from these tests only under the assumption that $\lambda=1$. Therefore, if our results from step one indicate that the gain score model is not a viable option, these tests cannot be viewed as reliable. Similarly, both estimators require strict exogeneity of the included regressors, and a violation of this assumption limits the reliability of this test.

Another approach to assessing the influence of unobserved heterogeneity that is applicable in both the gain score model and the lag score model is to investigate how observable child and family characteristics vary across different levels of exposure to particular teacher

¹¹ This is accomplished by estimating the gain score equation (7) with the addition of variables representing the within-child average values across time of all the time-varying variables. If a joint significance test of all the time averages rejects, it provides evidence that the random effects assumption is not justified.

characteristic (*TC*) or teaching practices (*PDG*). To do so, we adapt the common balance of covariates approach, in which the mean observable characteristics across levels in TC or PDG are compared. First we explore whether the variables in TC or PDG (which we can refer to as “treatments”) vary with the child-specific *observables*—prior achievement and the variables contained in X_{ijt} —in the following regression:¹²

$$(8) \quad \text{Teacher variable}_i = \alpha_t + \beta Y_{i,t-1} + \zeta X_{it} + v_{it}$$

If a test of joint significance of all the right-hand-side variables in (8)—i.e., a test of the null that all β and ζ are zero—does not reject, then there is strong evidence that treatment is unrelated to these observed child characteristics. In this case, we have more confidence that selection on *unobservables* is also negligible. If instead the test rejects, then we conclude that exposure to treatment differs systematically across child characteristics and, unless we believe that our set of observables is so complete as to eliminate anything that might be left in c_i (e.g., individual intelligence or motivation) which is correlated with treatment, we might continue to worry about unobserved heterogeneity.

If the test rejects, a possible reason might be that systematic variation of treatment across child and family characteristics is due to the sorting of children with particular characteristics into particular schools. To verify this, we can run the following regression, which includes both the child variables and school dummies, and again test the joint significance of the child variables:

$$(9) \quad \text{Teacher variable}_i = \alpha_t + \beta Y_{i,t-1} + \zeta X_{it} + \delta_s + v_{it}$$

If the inclusion of school dummies removes the significance of the child variables, then we can assume that treatments are distributed randomly across child characteristics *within*

¹² For computational feasibility, when the dependent variable is dichotomous, we use a linear probability model with discrete regressors.

schools. This would offer evidence that, at least within schools, selection on observables is negligible. If we again rely on the assumption that selection on the unobservables, c_i , should be no more threatening than that related to our rich set of observables, then we can argue that panel data methods that remove c_i are unnecessary, and that the inclusion of school dummies in either the cross-sectional or panel regressions will suffice to remove this threat. If instead, evidence of unobserved heterogeneity is non-negligible, then the panel will be needed at the cost of requiring the additional assumptions discussed in Section 2 for consistent estimation of the treatment effects.

Choosing an Estimator

The above analyses provide a strategy for choosing between the lag and the gain score model and deciding whether to pool or separate the data across grades. After gathering the information from these analyses, we can substantially narrow down the set of 18 choices illustrated in Figure 1. The gain score/lag score decision, coupled with the cross-section/panel decision will lead to a small number of estimators for further consideration. In addition, the assignment to treatment analysis helps determine whether or not to include school dummies. At this point, if more than one model/estimation method remains viable, we will compare their findings and discuss the similarities and differences found.

6. Results

Gain Score vs. Lag Score: Results

To choose between using the lag score equation (6) or the gain score equation (7) we first visually inspect the coefficient on prior test scores in (6) and then apply the test proposed by Harris, Sass, and Semykina (2011). When we use OLS to estimate (6) with the panel, our estimate of λ is 0.7505 and when we estimate it with the cross-sections, the estimate of λ is

0.9070 for kindergarten and 0.6998 for first grade. All three estimates are statistically different from 1, with the p-value equal to 0.0000 in each case for the test of the null hypothesis that $\lambda=1$. For the formal test, we estimate equation (7) and include the lagged inputs in the first grade specification. We find the lagged inputs to be statistically significant when included in equation (7), with a p-value for the joint test of significance of 0.0005. Given these results, we choose the lag score specification found in equation (6).

Cross Section vs. Panel: Results

To investigate whether teaching practices should be allowed to vary across grades, we estimated equation (6) using the pooled data and including interaction terms between the grade dummy and all the teacher practices and characteristics. The results are shown in Table 1. Due to the large number of controls included in the model, we only display those relevant to our research question – i.e., the teaching practices and teacher characteristics. We find significant interactions with grade for a small number of teacher characteristics and teaching practices. Post-graduate education matters more in predicting higher achievement in first grade than in kindergarten. Working with counting manipulatives predicts lower achievement in first grade than in kindergarten—outweighing the positive main effect. Overall, we find that pooling the data and constraining the coefficients to be the same across grades can obscure grade-specific relationships in a sufficient number of instances so as to make it helpful to separate our regressions by grade.

To assess whether there is evidence of selection on observables, we estimate equations (8) and (9)—i.e. the treatment regressions. The p-values for the joint F-test for these regressions are presented by grade in Table 2. For equation (8), our results show that for most teaching practices and most teacher characteristics we consider, the child variables are found to be jointly

significant, indicating that treatments are not randomly distributed across these observable characteristics. However, once we condition on the school attended by including school dummies in equation (9), the child variables are jointly significant in only one case—having a Hispanic teacher.¹³ In effect, all other teacher characteristics and instructional practices appear to be randomly assigned to students with different observable characteristics *within* schools. This offers evidence that the selection on observables is random within schools. Because of the rich set of observables these data provide, it can be argued that selection on unobservables should also be negligible. Thus, panel data methods that remove child effects (i.e., FE and AB) may be overly cautious—with the first differencing removing too much variation—if we already control for school differences with these data. Using AB instead of a simpler estimator with school dummies will likely increase standard errors due to the lost variation in the data and reduce our ability to detect significant effects. Furthermore, AB requires additional assumptions, such as no serial correlation in equation (6) and strict exogeneity of the other inputs.¹⁴ By eschewing AB, we do not rely on these assumptions to obtain consistent estimates.

As a result of these analyses, we narrow down the model/estimation choices in Figure 1 to those that use cross-section data (and thus preserve maximum flexibility) and include school dummies.

Note that previous studies relying on HLM techniques (e.g., Guarino et al. 2006, Bodovski & Farkas 2007, Parlady & Rumberger 2008, Rowan, Correnti, & Miller 2002) assumed random school effects, which are uncorrelated with the teacher characteristics and practices experienced by the child, rather than fixed school effects, which do not impose this

¹³ Further investigation (not shown in a table) revealed that some racial/ethnic matching may take place within schools, which led us to include interactions between teacher and child race categories on our achievement regressions.

¹⁴ AB also requires that the coefficient on the lagged dependent variable not be close to one. In that case, it breaks down due to a weak instrument problem.

assumption. The above analyses suggest that this assumption is unlikely to hold in ECLS-K, thus leading estimators with random school effects to be inconsistent.

Choosing an Estimator: Results

At this point, the remaining estimation choices among those outlined in Figure 1 are OLS and HLM on the lag score model with school dummy variables. Table 3 presents results for teacher characteristics and teaching practices using these estimators side by side. Certain teacher characteristics and practices—though relatively few—show evidence of producing achievement effects. The two methods produce very similar results.¹⁵ The main difference between the two approaches lies in the computation of the standard errors. We used cluster-robust standard errors at the school level in the OLS regressions, producing standard errors that tended to be slightly more conservative.

Teacher characteristics show different effects in kindergarten and first grade as expected due to our previous investigation of interaction terms. In kindergarten, we find evidence that having a teacher who is certified reduces achievement in kindergarten by approximately .43 IRT scale points (a small effect of about 1/25 of a standard deviation¹⁶) and that having taken more than two courses in methods of teaching mathematics reduces achievement by a little more than half that amount. In first grade, we find, however, that certification and coursework does not matter one way or the other, but that training in the form of advanced degrees has a positive effect: having a teacher with post-graduate education raise test scores by roughly four IRT scale points—i.e., more than one-third of a standard deviation. The differences in the effects of

¹⁵ In fact, the coefficients themselves are more or less identical for OLS and HLM in kindergarten due to the fact that the estimated unexplained variance in achievement due to differences in teachers *within* schools is very close to zero. In first grade, this variance component is somewhat larger.¹⁶ As shown in Appendix Table 1, the standard deviation in test scores is 8.84 in kindergarten and 8.65 in first grade.

¹⁶ As shown in Appendix Table 1, the standard deviation in test scores is 8.84 in kindergarten and 8.65 in first grade.

background training and education between kindergarten and first grade—discussed further in the next section—are noteworthy.

Small but for the most part strongly significant effects emerged in both grades with respect to teaching practices, and, again, the effects of specific practices differ across grades. In kindergarten, teachers who emphasize the use of counting manipulatives and the chalkboard have a positive impact on achievement. The effects are small—using any of these practices for an additional 10 days per month will raise test scores by less than $1/25$ of a standard deviation in IRT scale points. The smaller standard errors in the HLM regression also point to weak evidence that using math worksheets has a small positive effect and using creative movement to teach math has a small negative effect.

In first grade, teachers who spend more time engaging students in explaining how a mathematics problem is solved raise test scores by approximately .05 IRT scale points, and the coefficient is highly significant. Although somewhat larger than the other coefficients, it indicates that using this practice an extra 10 days per month will raise achievement by $1/20$ of a standard deviation—still a modest effect. A smaller and weakly significant positive effect is detected for doing mathematics problems from a textbook. In addition, the HLM regression provides some evidence that working on problems for which there are several solutions has a positive impact on achievement, a finding that complements the strong finding for explaining math problems.

Discussion

If we assert that the findings are causal, some policy implications emerge. Even though kindergarten and first grade represent closely spaced points on a continuum of early childhood education, we find that few characteristics and practices have a consistent effect in both grades.

Our results could be interpreted as suggesting that training and pedagogy that is geared toward analysis or explanation is appropriate for first grade but not necessarily for kindergarten. There are three possible explanations for this. One is that the ECLS tests measure very different constructs in the two grades. However, given that the tests are constructed by the same assessment teams so as to provide a certain amount of continuity, this seems unlikely. Or, secondly, curricular differences across the two grades may be such that they align more with standardized tests in first grade. However, given the large set of content coverage controls in our models, this also seems unlikely. A third explanation, that the cognitive development of a child differs markedly across these two periods of growth, is perhaps more likely. As a corollary to this hypothesis, the negative finding for certification in kindergarten warrants further investigation and suggests either that the approach to mathematics pedagogy in early elementary teacher trainings programs may not be geared toward achieving the kind of learning that can be measured by standardized tests administered at that point in time, or that it is not well aligned with learning development at the kindergarten stage. It is important to note that teacher training programs generally group kindergarten, early elementary, and upper elementary training into the same mathematical content and pedagogy courses. Thus training is not fine-tuned toward specific developmental stages.

Our findings with respect to the specific practices suggest that the developmental needs of students change from kindergarten to first grade and may require different teaching techniques. The use of counting manipulatives, although fairly frequently used in both grades, with kindergarten teachers using it an average of 12.5 times per month and first grade teachers using it an average of 11 times a month (see Appendix Table 1), only affected achievement in kindergarten. A possible explanation might be that the usefulness of this type of manipulative in

influencing learning reaches a plateau; kinesthetic approaches may prove effective among kindergarteners, whereas first grade students may outgrow their use. Also effective in kindergarten only is the use of chalkboards. Kindergarten teachers report using this practice only 4.7 times per month on average while first grade teachers use it almost twice that amount, suggesting that its relatively non-routine practice in kindergarten produces a helpful boost to achievement. On the other hand, the more verbally-oriented pedagogical technique of asking students to explain how a problem is solved influences achievement for first grade students but not for kindergarten students. It is encouraging to note that this practice is utilized relatively often in first grade (on average, 12.8 times per month in first grade versus 8 times per month in kindergarten).

Several explanations might be offered for the small pedagogical effects. First, it is important to acknowledge the limitations of retrospective survey items in accurately capturing the frequency of use. Although some studies have validated survey responses by comparing them with measurements taken during classroom observations (Mayer, 1999; Stipek & Byler, 2004), some measurement error and possibly recall bias may remain.¹⁷ Measurement error, if random, would attenuate coefficients. Another possible explanation is that pedagogical activities as isolated and specific as those we measure may have a small effect on learning and that several “best practices” or artful combinations of them are needed to produce substantial learning gains. Finally, it may be the case that it is not so much what a teacher does but how she does it that matters in producing learning. Thus measures of frequency and not quality of teaching modalities do not capture all the essential components of pedagogy. In other words, the effect of an effective kindergarten teacher who completes math problems on a chalkboard might differ from

¹⁷ That the “continuous” frequency scale was derived from more approximate, discrete frequency categories undoubtedly contributes to measurement error.

that of an ineffective one who does the same thing. In this sense, it should be acknowledged that despite the steps we have taken to select an estimation strategy, an impediment to claiming that our results are causal remains. The remaining issue is that neither the OLS nor the HLM estimator deals with the possibility that omitted teacher effects, represented in (6) and (7) by τ_j are nonrandom. As we have mentioned, one concern might be that high quality teachers tend to use certain teaching techniques but that such techniques, if adopted less skilled teachers, would produce little effect. Or, it is possible the seeming ineffectiveness of particular techniques may be due to inadequate training in those methods. Thus there is a question whether the technique itself or the ability of teachers to properly use the technique matters.

Our data follow a single cohort of students and do not provide the type of longitudinal information on their teachers that would permit us to control for time-constant teacher effects. We have no econometric technique at our disposition that allows us to eliminate unobserved teacher effects. We might argue, however, that the inclusion of the extensive set of teacher characteristics, content coverage variables, and school indicators, as well as our practices of interest, substantially narrows the range of what can be attributed to an unobserved teacher effect. Furthermore, in sensitivity analyses¹⁸ in which teacher characteristics are omitted from the model, the practice estimates change very little, suggesting that their use is not driven by training or experience in ways that affect achievement.

Sensitivity of Findings to Other Specifications and Estimators

As mentioned in our review of the literature, prior survey-based studies have tended to use different models and estimators from those selected in our analysis. In order to demonstrate the importance of taking careful steps in selecting the methods used, Table 4 shows the sensitivity of findings to the choice of model and estimator. The table displays results for HLM

¹⁸ These analyses are not shown in the paper but are available from the authors upon request.

estimation with random school effects on the kindergarten and first grade cross-sections in the first four columns—analytic approaches that have been more prevalent in the literature than our preferred approaches—and results for the child fixed effects estimator applied to the panel in the last column. The HLM estimation is carried out for both the lag score and the gain score model.

Before discussing the differences between these approaches and our preferred approaches, it is interesting to note the differences between the lag and gain score models in the results displayed here in Table 4. Moving to a gain score model has little effect on the magnitude of the estimates for kindergarten but a noticeable effect on those for first grade. Recall that in these data, the coefficient on λ was fairly close to one for kindergarten and much lower for first grade. In addition, our test of the $\lambda=1$ assumption applied only to first grade. Thus it appears that using the gain score model in the ECLS-K data is relatively costless in kindergarten, but these results serve to support our claim that doing so in first grade introduces a noticeable amount of bias.

Given this evidence, it seems all the more unlikely that the FE estimator in column five has much insight to offer, since it not only relies on a gain score model but also constrains the coefficients on teacher characteristics and practices to be the same across grades and relies on the strict exogeneity assumption. It is also interesting to note that the FE estimator, having removed a great deal of variation by time-demeaning the data, displays larger standard errors and thus finds no variable to be significant at the .05 level. For instance, the point estimate for the effect of having a teacher with regular certification is of similar or greater magnitude to the others; however the standard error is approximately twice as large. Differences in significance due to large changes in the point estimates (for example, see the coefficient on “very experienced”), on the other hand, may be due to bias—for example due to violations of strict exogeneity. Overall,

the FE estimator, applied to these data and research questions, offers little in the way of policy-relevant information with regard to our research question.

Moving now to a comparison of results from the HLM estimators containing random school effects (i.e., those commonly used in the literature) with those derived from our preferred estimators, shown in Table 3, we find a few similarities but some notable differences. With regard to teacher characteristics, the HLM results from Table 4 generally display coefficients with slightly lower magnitudes for certification and coursework in kindergarten and coefficients with more notably lower magnitudes for postgraduate study in first grade.

With regard to instructional practices, the Table 4 HLM estimates diverge from those in Table 3 in several instances. In kindergarten, they show a significant negative effect of counting out loud, the use of geometric manipulatives, engaging in calendar-related activities, whereas these are not significant when school effects are treated as fixed. In addition, although both sets of results display significant coefficients for working with counting manipulatives and completing problems on the chalkboard, the magnitudes differ slightly. In first grade, the magnitude and significance of counting out loud, the use of calendars, and working on problems with several solutions differs across the two sets of results, and the magnitude of the most robust results—that of explaining how problems are solved—is somewhat understated.

It is important to reiterate why we might see the differences outlined above. Differences in the conclusions based on different estimators can be driven either by differences in the point estimates or in the standard errors. Of the two HLM approaches, the school random effects specification tends to produce smaller standard errors due to the additional covariance structure imposed on the estimation. However, it is not the case that differences in the conclusions drawn between the two HLM estimators are purely due to smaller standard errors. Rather, it is typically

the case that the magnitude of the point estimates changes enough to alter the significance level. Such differences in the point estimates likely occur because estimates with school random effects do not properly control for the sorting of students into different schools, whereas when the random school effects are replaced by fixed effects, this sorting is explicitly accounted for.¹⁹

It is important to note that not only our preferred approaches but also our sensitivity analyses using HLM with random school effects produce some results that differ from the HLM studies in the prior literature and lead to different conclusions, particularly for first grade. This happens despite our attempts to mimic their methods in our sensitivity analyses. In these cases, specification and sample differences account for much of the divergence.

It is difficult to compare our findings with those of the two kindergarten studies we cited (Guarino et al., 2006 and Bodovski and Farkas, 2007), because the prior studies used instructional practice scales rather than individual practice variables. Both prior studies found that “traditional approaches” were positive and significantly related to achievement and thus conflated the effects of some of the variables we use (i.e., worksheets, textbooks, and chalkboard), one of which we find highly significant. However, our findings did not concur with other findings of theirs, such as the positive impact of student centered instruction in Guarino et al. (a scale composed of items such as explaining how mathematics problems are solved, playing games, and using music or creative movement), and interactive instruction in Bodovski and Farkas (a scale composed of items such as explaining how mathematics problems are solved, solving problems in small groups or with a partner, and peer tutoring).

The differences between our results and those of Parlady and Rumberger’s 2008 study of first grade are more stark and seemingly due to sample differences. The prior study used the 30 percent subsample of first graders who were sampled in the fall as well as the spring and did not

¹⁹ In other words, by controlling explicitly for δ_s in equations (6) and (7), rather than assuming it is random.

impute missing data. Their findings that the use of math worksheets and calendars raise and the use of geometric manipulatives lower first grade achievement are not supported in our study based on the full sample, even when we employ methods that are similar to theirs.

7. Summary and Conclusions

Survey data on instructional practices, if well designed and carefully analyzed, have the potential to address questions regarding the means by which effective teachers can affect student success. Simple teacher performance studies based on administrative data can be useful in showing that quality matters but have access to very limited information on teachers and provide little insight regarding policy prescriptions needed to improve overall effectiveness. Surveys that use richly detailed survey data to focus on teachers' actions and how they affect student outcomes can hold the key to designing policy instruments, assuming the results they find are causal.

A source of concern in non-experimental research on teacher effects, however, is the sensitivity of findings to different modeling and estimation techniques. This paper has outlined a process for selecting an appropriate model and estimation method to investigate teacher effects on student achievement using longitudinal survey data. We applied this process in a series of sequential steps to data from ECLS-K and found little evidence to support many of the modeling and estimation choices heavily relied upon in the literature—for example, models with gain scores or random school effects or models that constrain coefficients to be the same across grades.

Our study clearly illustrates how methodological choices can influence results. Although a few of our findings concur with those in prior research, most diverge. Prior studies find little or no relationship between teacher background characteristics and mathematics achievement in

either kindergarten or first grade, but we find evidence that teacher certification may slightly lower achievement in kindergarten and that postgraduate education may contribute to relatively substantial achievement gains in first grade. In addition, whereas prior studies claim that student-centered or interactive pedagogy improves achievement in kindergarten and that worksheets and calendars are important tools in first grade, we find that working with counting manipulatives and completing math problems on the chalkboard improve achievement in kindergarten and that explaining how mathematics problems are solved is important in first grade. Taken as a whole, our findings suggest that there may be important developmental differences in the mathematics learning capabilities of children in kindergarten versus first grade and that training and pedagogy should be structured appropriately.

Importantly, there is no finding that holds up under all the many specification and estimation approaches available to researchers—a circumstance that highlights the need to ground the selection of a model and estimation method on sound reasoning. Clearly, it is important that researchers justify their methodological choices through a thorough investigation of related assumptions and sensitivities. The guidelines for empirical investigation in the longitudinal survey data context provided in this study can aid researchers in selecting a credible set of results and advancing causal claims.

References

- Arellano, M. & Bond, S. (1991) Some Tests of Specification of Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58, pp.277-298.
- Barnett, S.W. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5, 25-50.
- Bodovski, K. & Farkas, G. (2007). Do Instructional Practices Contribute to Inequality in Achievement? The Case of Mathematics Instruction in Kindergarten. *Journal of Early Childhood Research*, 5(3), 301-322.
- Carlin, J. B., Galati, J. C., & Royston, P. (2008). A New Framework for Managing and Analyzing Multiply Imputed Data in Stata. *Stata Journal*, 8(1), 49-67.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Whitmore Schanzenbach, D., & Yagan, D. (2010) How does your kindergarten classroom affect your earnings? Evidence from Project STAR. NBER working paper 16381.
- Cohen, D. & Hill, H. (2000). Instructional Policy and Classroom Performance: the Mathematics Reform in California. *Teachers College Record*, 102(2): 294-343.
- Currie, J., Thomas, D. (2000). School quality and the longer-term effects of Head Start. *The Journal of Human Resources*, 35, 4, 755-774.
- Guarino, C., Hamilton, L., Lockwood, J.R., & Rathbun, A.H. (2006). *Teacher Qualifications, Instructional Practices, and Reading and Mathematics Gains of Kindergarteners*. (NCES 2006-031). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Guarino, C., Reckase, M. & Wooldridge, J. (unpublished draft). Evaluating Value-Added Methods for Estimating Teacher Effects.

Hamilton, L., McCaffrey, D., Stecher, B., Klein, S., Robyn, A., & Bugliari, D. (2003). Studying Large-Scale Reforms of instructional Practice: An Example from Mathematics and Science. *Educational Evaluation and Policy Analysis*, 25(1): 1-29.

Hanushek, E. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, 14(3): 351-388, 1979.

Harris, D., Sass, T., & Semykina, A. (unpublished draft). Value-Added Models and the Measurement of Teacher Productivity.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46, 6, 1251-1271.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it Up: Helping Children Learn Mathematics*. National Academy Press, Washington D.C.

Le, V., Stecher, B., Lockwood, J.R., Hamilton, L., Robyn, A., Williams, V., et al. (2006). Improving mathematics and science education: A longitudinal investigation between reform-oriented instruction and student achievement. RAND Corporation, Santa Monica, CA.

Mayer, D.P. (1999). Measuring instructional practice: Can policy makers trust survey data? *Educational Evaluation and Policy Analysis*, 21, 1, 29-45.

Palardy, G. & Rumberger, R. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, 30(2), 111-140.

Rock, D.A., & Pollack, J.M. (2002). *Early Childhood Longitudinal Study - Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report for Kindergarten Through First Grade*

- (NCES 2002–005). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Rowan, B., Correnti, R., & Miller, R. (2002). What Large-Scale Survey Research tells us about Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record*, 104, 1525–67.
- Royston, P. (2004). Multiple Imputation of Missing Values. *Stata Journal*, 4(3), 227-241.
- Stipek, D., Byler, P. (2004) The early childhood classroom observation measure. *Early Childhood Research Quarterly*, 19, 375-397.
- Todd, P. & Wolpin, K. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *Economic Journal*, 113(485), 3-33.
- Tourangeau, K., Burke, J., Le, T., Wan, S., Weant, M., Brown, E., Vaden-Kiernan, N., Rinker, E., Dulaney, R., Ellingson, K., Barrett, B., Flores-Cervantes, I., Zill, N., Pollack, J., Rock, D., Atkins-Burnett, S., Meisels, S., Bose, J., West, J., Denton, K., Rathbun, A., & Walston, J. (2001). *ECLS-K Base Year Public-use Data Files and Electronic Codebook: User's Manual* .(NCES 2001–029). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Van Buuren, S., Boshuizen, H., & Knock, D. (1999). Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. *Statistics in Medicine*, 18:681-694.
- Wayne, A. & Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research*, 73, 1, 89-122.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, Ma.

Table 1: OLS Estimates of Lag Score Specification Pooled with Grade Interactions with School Dummies

	Base	1 st Grade Interaction
Teacher Characteristics		
<i>Black</i>	-0.3012 (0.5627)	-0.4136 (0.7835)
<i>Other</i>	-0.1369 (0.4828)	0.0286 (0.6619)
<i>Hispanic</i>	-0.4883 (0.5501)	0.7102 (0.8390)
<i>Novice</i>	-0.1978 (0.2303)	0.1326 (0.3360)
<i>Very experienced</i>	-0.5538*** (0.2018)	0.4035 (0.3001)
<i>Above BA</i>	0.1938 (0.1999)	1.9183** (0.8159)
<i>Masters or above</i>	0.1571 (0.2151)	2.1143** (0.8216)
<i>Regular certification</i>	-0.3349 (0.2591)	-0.1044 (0.3528)
<i>Courses in methods for teaching math</i>	0.0142 (0.1567)	0.0365 (0.2191)
<i>Paid preparatory time</i>	0.0514 (0.1994)	-0.0789 (0.2572)
<i>Non-paid preparatory time</i>	0.1030 (0.1525)	-0.0463 (0.1994)
Pedagogy Practices		
<i>Hours per week teaching math</i>	0.0208 (0.0408)	0.0073 (0.0648)
<i>Hours per week in math groups</i>	-0.0633 (0.0765)	0.0486 (0.1009)
<i>Count out loud</i>	-0.0341* (0.0202)	0.0132 (0.0239)
<i>Work with geometric manipulatives</i>	-0.0333** (0.0139)	0.0379* (0.0224)
<i>Work with counting manipulatives to learn basic operations</i>	0.0544*** (0.0175)	-0.0680*** (0.0221)
<i>Play mathematics-related games</i>	-0.0220 (0.0145)	0.0140 (0.0201)
<i>Use a calculator for math</i>	0.0219 (0.0327)	0.0029 (0.0476)
<i>Use music to understand mathematics concepts</i>	-0.0016 (0.0144)	-0.0319 (0.0313)
<i>Use creative movement or drama to understand mathematics concepts</i>	-0.0226 (0.0189)	-0.0012 (0.0315)

What can we learn about effective early mathematics teaching?

<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	-0.0153 (0.0202)	-0.0284 (0.0298)
<i>Explain how a mathematics problem is solved</i>	0.0144 (0.0128)	0.0181 (0.0183)
<i>Engage in calendar-related activities</i>	0.0162 (0.0244)	-0.0069 (0.0285)
<i>Do mathematics worksheets</i>	0.0217 (0.0133)	-0.0075 (0.0181)
<i>Do mathematics problems from the textbook</i>	0.0297** (0.0135)	-0.0082 (0.0162)
<i>Complete mathematics problems on the chalkboard</i>	0.0359** (0.0141)	-0.0202 (0.0180)
<i>Solve mathematics problems in small groups or with a partner</i>	0.0001 (0.0154)	0.0139 (0.0224)
<i>Work on mathematics problems that reflect real-life situations</i>	0.0043 (0.0149)	0.0045 (0.0208)
<i>Work in mixed achievement groups</i>	-0.0074 (0.0113)	0.0052 (0.0155)
<i>Peer tutoring</i>	-0.0037 (0.0121)	0.0118 (0.0180)

Observations

21,232

Standard errors clustered at the school level in parentheses; estimation with 40 imputed data sets

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Random Assignment Test p-values: Joint Significance of Child Characteristics and Lagged Score

Dependent Variable	Kindergarten		1 st Grade	
	Without School Dummies	With School Dummies	Without School Dummies	With School Dummies
Pedagogy Practices				
<i>Hours per week teaching math</i>	0.0000	0.6841	0.1145	0.9220
<i>Hours per week in math groups</i>	0.0000	0.1051	0.0000	0.9683
<i>Count out loud</i>	0.0011	0.2916	0.0000	0.8087
<i>Work with geometric manipulatives</i>	0.0000	0.4642	0.0000	0.9678
<i>Work with counting manipulatives to learn basic operations</i>	0.0003	0.2363	0.0004	0.3322
<i>Play mathematics-related games</i>	0.0928	0.1510	0.3604	0.8725
<i>Use a calculator for math</i>	0.3464	0.9923	0.0001	0.9976
<i>Use music to understand mathematics concepts</i>	0.0008	0.2787	0.0136	0.9748
<i>Use creative movement or drama to understand mathematics concepts</i>	0.0000	0.6779	0.0000	0.8972
<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	0.0046	0.2788	0.2348	0.2008
<i>Explain how a mathematics problem is solved</i>	0.0000	0.7656	0.2025	0.2329
<i>Engage in calendar-related activities</i>	0.0170	0.3226	0.0002	0.9661
<i>Do mathematics worksheets</i>	0.0000	0.9823	0.2742	0.6238
<i>Do mathematics problems from the textbook</i>	0.0000	0.9489	0.0000	0.8261
<i>Complete mathematics problems on the chalkboard</i>	0.0000	0.9054	0.0000	0.9381
<i>Solve mathematics problems in small groups or with a partner</i>	0.0000	0.9257	0.0048	0.8505
<i>Work on mathematics problems that reflect real-life situations</i>	0.0200	0.9708	0.0209	0.9313
<i>Work in mixed achievement groups</i>	0.0033	0.3245	0.0000	0.7647
<i>Peer tutoring</i>	0.0037	0.6531	0.0000	0.9711
<i>Do worksheets or workbook page emphasizing routing practice or drill</i>			0.0025	0.9871
<i>Work on problems for which there are several solutions</i>			0.0000	0.6069
Teacher Characteristics				
<i>Black</i>	0.0000	0.4198	0.0000	0.7903
<i>Other</i>	0.0017	0.8295	0.0004	0.5561
<i>Hispanic</i>	0.0000	0.0002	0.0000	0.0105
<i>Novice</i>	0.0762	0.5053	0.0014	0.9622
<i>Very experienced</i>	0.0026	0.3906	0.0012	0.9528
<i>Above BA</i>	0.0000	0.8872	0.0001	0.7972
<i>Masters or above</i>	0.0000	0.8701	0.0000	0.7798
<i>Regular certification</i>	0.0000	0.4256	0.0000	0.9374
<i>Courses in methods for teaching math</i>	0.2598	0.5452	0.0660	0.8519
<i>Paid preparatory time</i>	0.0000	0.9813	0.0000	0.9950

What can we learn about effective early mathematics teaching?

Non-paid preparatory time

0.0004

0.6177

0.0089

0.9681

Table 3: Main Model and Estimation Results

Estimation Method	Estimation and Modeling Choices			
	<i>OLS</i>	<i>MLE/HLM</i>	<i>OLS</i>	<i>MLE/HLM</i>
Specification	<i>Lag</i>	<i>Lag</i>	<i>Lag</i>	<i>Lag</i>
School Effect	<i>Fixed</i>	<i>Fixed</i>	<i>Fixed</i>	<i>Fixed</i>
Teacher Effect	<i>None</i>	<i>Random</i>	<i>None</i>	<i>Random</i>
Grades	<i>Kindergarten</i>	<i>Kindergarten</i>	<i>1st Grade</i>	<i>1st Grade</i>
Teacher Characteristics				
<i>Black</i>	0.1236 (0.4491)	0.1236 (0.3879)	-0.8861 (0.6216)	-0.5563 (0.4933)
<i>Other</i>	-0.0395 (0.3962)	-0.0395 (0.3331)	0.3187 (0.5209)	0.453 (0.3942)
<i>Hispanic</i>	-0.0002 (0.4823)	-0.0002 (0.4108)	0.245 (0.7377)	0.6662 (0.6330)
<i>Novice</i>	-0.1474 (0.2128)	-0.1474 (0.1767)	0.3012 (0.2790)	0.2489 (0.2252)
<i>Very experienced</i>	-0.1395 (0.1897)	-0.1395 (0.1499)	-0.0922 (0.2383)	-0.0997 (0.1957)
<i>Above BA</i>	-0.1015 (0.1813)	-0.1015 (0.1641)	4.066** (2.0097)	3.7029** (1.5919)
<i>Masters or above</i>	0.1616 (0.1871)	0.1616 (0.1635)	4.1721** (1.9963)	3.8633** (1.5822)
<i>Regular certification</i>	-0.4326* (0.2462)	-0.4326** (0.2042)	-0.209 (0.3037)	-0.2528 (0.2489)
<i>Courses in methods for teaching math</i>	-0.2555* (0.1434)	-0.2555** (0.1261)	0.0069 (0.1879)	-0.0291 (0.1538)
<i>Paid preparatory time</i>	0.0041 (0.1862)	0.0041 (0.1556)	0.0072 (0.2297)	-0.0333 (0.1956)
<i>Non-paid preparatory time</i>	0.1785 (0.1551)	0.1785 (0.1250)	0.1188 (0.2054)	0.1176 (0.1595)
Pedagogy Practices				
<i>Hours per week teaching math</i>	0.0138 (0.0385)	0.0138 (0.0327)	0.0171 (0.0657)	0.012 (0.0594)
<i>Hours per week in math groups</i>	-0.0114 (0.0722)	-0.0114 (0.0617)	-0.0733 (0.0786)	-0.0799 (0.0615)
<i>Count out loud</i>	-0.0247 (0.0189)	-0.0247 (0.0156)	-0.024 (0.0166)	-0.0274** (0.0133)
<i>Work with geometric manipulatives</i>	-0.0096 (0.0141)	-0.0096 (0.0112)	0.0084 (0.0192)	0.0081 (0.0153)
<i>Work with counting manipulatives to learn basic operations</i>	0.0323** (0.0152)	0.0323** (0.0129)	-0.0162 (0.0180)	-0.0174 (0.0150)
<i>Play mathematics-related games</i>	-0.017 (0.0122)	-0.017 (0.0107)	-0.0152 (0.0183)	-0.0107 (0.0155)

<i>Use a calculator for math</i>	-0.0021 (0.0299)	-0.0021 (0.0240)	-0.0016 (0.0358)	-0.0058 (0.0304)
<i>Use music to understand mathematics concepts</i>	0.0049 (0.0127)	0.0049 (0.0109)	-0.0355 (0.0289)	-0.0306 (0.0236)
<i>Use creative movement or drama to understand mathematics concepts</i>	-0.0264 (0.0163)	-0.0264* (0.0136)	0.0091 (0.0271)	0.0141 (0.0235)
<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	-0.0018 (0.0177)	-0.0018 (0.0149)	-0.027 (0.0274)	-0.0213 (0.0217)
<i>Explain how a mathematics problem is solved</i>	-0.0026 (0.0114)	-0.0026 (0.0101)	0.0494*** (0.0165)	0.0489*** (0.0130)
<i>Engage in calendar-related activities</i>	0.0123 (0.0237)	0.0123 (0.0204)	-0.0027 (0.0196)	-0.0023 (0.0162)
<i>Do mathematics worksheets</i>	0.0206 (0.0127)	0.0206* (0.0109)	0.0078 (0.0177)	0.0066 (0.0144)
<i>Do mathematics problems from the textbook</i>	0.0088 (0.0145)	0.0088 (0.0127)	0.0275* (0.0143)	0.0272** (0.0115)
<i>Complete mathematics problems on the chalkboard</i>	0.0280** (0.0127)	0.0280** (0.0117)	0.0172 (0.0142)	0.0174 (0.0113)
<i>Solve mathematics problems in small groups or with a partner</i>	-0.0055 (0.0141)	-0.0055 (0.0117)	0.0114 (0.0201)	0.0084 (0.0157)
<i>Work on mathematics problems that reflect real-life situations</i>	0.0003 (0.0128)	0.0003 (0.0106)	-0.0211 (0.0177)	-0.0217 (0.0143)
<i>Work in mixed achievement groups</i>	0.0063 (0.0106)	0.0063 (0.0086)	0.0102 (0.0141)	0.0123 (0.0115)
<i>Peer tutoring</i>	-0.0096 (0.0118)	-0.0096 (0.0096)	0.0164 (0.0159)	0.0187 (0.0135)
<i>Do worksheets or workbook page emphasizing routing practice or drill</i>			0.005 (0.0173)	0.0021 (0.0141)
<i>Work on problems for which there are several solutions</i>			0.0237 (0.0171)	0.0286** (0.0141)
Observations	16,356	16,356	11,780	11,780

Source: ECLS-K; Standard errors clustered at the school level in parentheses; estimation with 40 imputed data sets

*** p<0.01, ** p<0.05, * p<0.1

All specifications include class size, class racial percentages, teacher content practices, family welfare status, prior test score, time between exams, household size, number of siblings, level of parental involvement, how often the child reads, participation in extracurricular activities, number of children's books; indicators for completing the Spanish math exam, student disability, mother's education level, household income level, mother's employment status, single parent family, father absent, pay tuition, child's race, speak non

Table 4: Alternative Models and Estimators

Estimation Method	Estimation and Modeling Choices				
	MLE/HLM	MLE/HLM	MLE/HLM	MLE/HLM	FE
Specification	<i>Lag</i>	<i>Gain</i>	<i>Lag</i>	<i>Gain</i>	<i>Gain</i>
School Effect	<i>Random</i>	<i>Random</i>	<i>Random</i>	<i>Random</i>	<i>None</i>
Teacher Effect	<i>Random</i>	<i>Random</i>	<i>Random</i>	<i>Random</i>	<i>None</i>
Grades	<i>Kindergarten</i>	<i>Kindergarten</i>	<i>1st Grade</i>	<i>1st Grade</i>	<i>Pooled</i>
Teacher Characteristics					
<i>Black</i>	0.0820 (0.4290)	0.1019 (0.4302)	-0.7731 (0.5019)	-0.6880 (0.5146)	-0.9328 (0.9257)
<i>Other</i>	-0.3261 (0.3042)	-0.2909 (0.3024)	0.0946 (0.4084)	0.3276 (0.4344)	-0.1778 (0.8656)
<i>Hispanic</i>	-0.2503 (0.4181)	-0.1968 (0.4222)	0.8086 (0.6257)	0.6579 (0.6573)	-0.2694 (0.9521)
<i>Novice</i>	-0.2577 (0.1625)	-0.2569 (0.1626)	0.0533 (0.2289)	0.0218 (0.2414)	-0.2632 (0.3565)
<i>Very experienced</i>	-0.1416 (0.1363)	-0.1616 (0.1358)	-0.0957 (0.1955)	-0.1045 (0.2073)	-0.5128* (0.2958)
<i>Above BA</i>	-0.1562 (0.1420)	-0.1394 (0.1411)	2.3174*** (0.7260)	2.5012*** (0.7687)	-0.3458 (0.4092)
<i>Masters or above</i>	0.0017 (0.1473)	0.0126 (0.1474)	2.5502*** (0.7256)	2.8267*** (0.7685)	0.0018 (0.4388)
<i>Regular certification</i>	-0.4011** (0.1730)	-0.3845** (0.1728)	-0.2826 (0.2399)	-0.3401 (0.2483)	-0.4328 (0.4258)
<i>Courses in methods for teaching math</i>	-0.1980* (0.1145)	-0.1772 (0.1136)	-0.2060 (0.1514)	-0.1873 (0.1617)	0.0441 (0.2566)
<i>Paid preparatory time</i>	-0.0333 (0.1279)	-0.0458 (0.1282)	-0.1171 (0.1713)	-0.1895 (0.1856)	-0.0019 (0.3024)
<i>Non-paid preparatory time</i>	0.1539 (0.1086)	0.1520 (0.1091)	0.1777 (0.1499)	0.1018 (0.1582)	0.0746 (0.2534)
Pedagogy Practices					
<i>Hours per week teaching math</i>	0.0419 (0.0289)	0.0462 (0.0288)	0.0647 (0.0564)	0.0349 (0.0587)	0.0973 (0.0694)
<i>Hours per week in math groups</i>	0.0271 (0.0584)	0.0241 (0.0587)	-0.0460 (0.0643)	-0.0151 (0.0657)	-0.0225 (0.1082)
<i>Count out loud</i>	-0.0337** (0.0136)	-0.0309** (0.0137)	-0.0183 (0.0121)	-0.0011 (0.0132)	-0.0354 (0.0245)
<i>Work with geometric manipulatives</i>	-0.0185* (0.0096)	-0.0181* (0.0097)	0.0036 (0.0149)	0.0072 (0.0154)	-0.0335 (0.0228)
<i>Work with counting manipulatives to learn basic operations</i>	0.0253** (0.0113)	0.0274** (0.0112)	-0.0181 (0.0136)	-0.0095 (0.0148)	0.0446* (0.0258)
<i>Play mathematics-related games</i>	-0.0170* (0.0092)	-0.0191** (0.0093)	-0.0052 (0.0149)	0.0029 (0.0157)	-0.0084 (0.0217)
<i>Use a calculator for math</i>	-0.0027 (0.0225)	-0.0040 (0.0224)	0.0090 (0.0304)	-0.0014 (0.0313)	0.0550 (0.0517)

What can we learn about effective early mathematics teaching?

<i>Use music to understand mathematics concepts</i>	0.0152 (0.0100)	0.0171* (0.0100)	-0.0102 (0.0265)	0.0042 (0.0279)	-0.0005 (0.0280)
<i>Use creative movement or drama to understand mathematics concepts</i>	-0.0255* (0.0132)	-0.0264** (0.0133)	-0.0144 (0.0248)	-0.0155 (0.0255)	-0.0479 (0.0356)
<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	-0.0085 (0.0133)	-0.0102 (0.0132)	-0.0060 (0.0204)	-0.0025 (0.0216)	-0.0330 (0.0302)
<i>Explain how a mathematics problem is solved</i>	-0.0016 (0.0085)	-0.0038 (0.0085)	0.0372*** (0.0122)	0.0288** (0.0127)	0.0059 (0.0208)
<i>Engage in calendar-related activities</i>	0.0311* (0.0158)	0.0314** (0.0158)	0.0239* (0.0138)	0.0221 (0.0144)	0.0120 (0.0300)
<i>Do mathematics worksheets</i>	0.0240*** (0.0082)	0.0243*** (0.0082)	0.0197 (0.0127)	0.0209 (0.0137)	0.0290 (0.0203)
<i>Do mathematics problems from the textbook</i>	-0.0028 (0.0090)	-0.0027 (0.0090)	0.0331*** (0.0096)	0.0349*** (0.0100)	0.0240 (0.0178)
<i>Complete mathematics problems on the chalkboard</i>	0.0309*** (0.0094)	0.0312*** (0.0094)	0.0123 (0.0110)	0.0133 (0.0118)	0.0320 (0.0202)
<i>Solve mathematics problems in small groups or with a partner</i>	0.0106 (0.0111)	0.0115 (0.0111)	-0.0102 (0.0147)	-0.0113 (0.0155)	0.0066 (0.0234)
<i>Work on mathematics problems that reflect real-life situations</i>	0.0029 (0.0097)	0.0032 (0.0097)	-0.0009 (0.0134)	-0.0039 (0.0143)	0.0232 (0.0221)
<i>Work in mixed achievement groups</i>	-0.0027 (0.0077)	-0.0030 (0.0077)	0.0043 (0.0112)	0.0038 (0.0119)	-0.0082 (0.0178)
<i>Peer tutoring</i>	-0.0058 (0.0087)	-0.0040 (0.0087)	0.0122 (0.0130)	0.0146 (0.0134)	-0.0050 (0.0189)
<i>Do worksheets or workbook page emphasizing routing practice or drill</i>			0.0069 (0.0128)	0.0113 (0.0137)	
<i>Work on problems for which there are several solutions</i>			0.0204 (0.0134)	0.011 (0.0140)	
Observations	16,356	16,356	11,780	11,780	21,232

Source: ECLS-K; Standard errors clustered at the school level in parentheses; estimation with 40 imputed data sets

*** p<0.01, ** p<0.05, * p<0.1

All specifications include class size, class racial percentages, teacher content practices, family welfare status, prior test score, time between exams, household size, number of siblings, level of parental involvement, how often the child reads, participation in extracurricular activities, number of children's books; indicators for completing the Spanish math exam, student disability, mother's education level, household income level, mother's employment status, single parent family, father absent, pay tuition, child's race, speak non-English in home, repeating kindergarten, attend full day kindergarten. Specifications without school dummies also include school level variables: minority percentage, private religious, private non-religious, school enrollment, region, suburban, rural, gang problems, crime problems.

Appendix Table 1: Summary Statistics for Forty Imputed Data Sets

	Pooled		Kindergarten		1 st Grade	
	Mean	SD	Mean	SD	Mean	SD
Child and Family Variables						
<i>Math score</i>	36.41	11.67	27.71	8.84	44.08	8.65
<i>Female</i>	0.498	0.500	0.493	0.500	0.495	0.500
<i>Non-English at home</i>	0.111	0.314	0.122	0.327	0.126	0.331
<i>Age in months</i>	80.95	7.383	74.67	4.44	86.97	4.21
<i>Household size</i>	4.722	1.781	2.077	0.268	8.942	6.415
<i>Number of siblings</i>	1.487	1.116	1.406	0.507	1.423	0.510
<i>Mother work before KG</i>	0.747	0.435	0.746	0.435	0.741	0.438
<i>How often child reads</i>	4.649	2.339	4.439	2.513	4.811	2.182
<i>Children's books in home</i>	82.76	61.58	74.41	59.31	85.11	62.46
<i>Welfare participant</i>	0.076	0.266	0.128	0.335	0.055	0.229
<i>Pay tuition</i>	0.210	0.407	0.213	0.409	0.209	0.407
<i>Spanish exam</i>	0.022	0.147	0.032	0.177	0.016	0.127
<i>Full day KG</i>	0.559	0.496	0.558	0.497	0.569	0.495
<i>Disabled</i>	0.142	0.349	0.140	0.347	0.158	0.365
<i>Single parent household</i>	0.200	0.400	0.223	0.416	0.241	0.428
<i>Father absent</i>	0.207	0.405	0.227	0.419	0.242	0.428
<i>Days between exams</i>	275.3	90.82	186.3	21.26	364.2	20.14
<i>KG repeater</i>	0.041	0.198	0.045	0.208	0.042	0.201
<i>Child Black</i>	0.132	0.338	0.150	0.357	0.136	0.343
<i>Child Hispanic</i>	0.161	0.368	0.173	0.379	0.162	0.368
<i>Child Asian or Pacific-Islander</i>	0.057	0.232	0.059	0.235	0.060	0.237
<i>Child other race</i>	0.044	0.205	0.044	0.205	0.043	0.204
<i>Mother below high school</i>	0.116	0.320	0.134	0.340	0.128	0.334
<i>Mother high school</i>	0.300	0.458	0.305	0.460	0.296	0.457
<i>Mother some college</i>	0.336	0.472	0.326	0.469	0.332	0.471
<i>Mother BA or above</i>	0.083	0.276	0.079	0.269	0.083	0.275
<i>Only Child</i>	0.146	0.354	0.001	0.023	0.001	0.023
<i>Mother work full time</i>	0.470	0.499	0.462	0.499	0.481	0.500
<i>Mother work part time</i>	0.229	0.420	0.217	0.413	0.223	0.416
<i>Parental school involvement</i>	3.973	1.522	3.679	1.576	4.035	1.539
<i>Extracurricular activities</i>	1.328	1.258	1.056	1.138	1.489	1.312
<i>Income less than \$20k</i>	0.203	0.402	0.239	0.427	0.226	0.418
<i>Income more than \$100k</i>	0.220	0.414	0.191	0.393	0.223	0.416
Teacher Characteristics						
<i>Teacher Black</i>	0.062	0.242	0.065	0.246	0.066	0.248
<i>Teacher Hispanic</i>	0.057	0.232	0.063	0.243	0.057	0.232
<i>Teacher other race</i>	0.050	0.217	0.048	0.213	0.056	0.230
<i>More than two course in teaching math</i>	0.445	0.497	0.426	0.495	0.461	0.499
<i>Less than five years experience</i>	0.199	0.399	0.184	0.388	0.228	0.420
<i>More than nine years experience</i>	0.621	0.485	0.618	0.486	0.600	0.490
<i>Certification status</i>	0.874	0.331	0.858	0.349	0.876	0.329
<i>Degree greater than BA</i>	0.323	0.467	0.349	0.477	0.298	0.457

What can we learn about effective early mathematics teaching?

<i>Degree MA or above</i>	0.523	0.499	0.353	0.478	0.692	0.461
<i>More than two hours paid preparatory time</i>	0.715	0.451	0.710	0.454	0.705	0.456
<i>More than five hours non-paid preparatory time</i>	0.431	0.495	0.406	0.491	0.444	0.497
Teacher Pedagogy Practices						
<i>Count out loud</i>	15.49	6.256	17.91	4.220	13.31	6.942
<i>Work with geometric manipulatives</i>	7.624	6.237	9.616	6.484	5.918	5.511
<i>Work with counting manipulatives to learn basic operations</i>	11.63	6.284	12.514	6.087	11.096	6.474
<i>Play mathematics-related games</i>	9.508	6.344	10.623	6.551	8.376	6.033
<i>Use a calculator for math</i>	0.952	2.424	0.609	2.296	1.310	2.548
<i>Use music to understand mathematics concepts</i>	2.692	4.719	4.045	5.607	1.461	3.243
<i>Use creative movement or drama to understand mathematics concepts</i>	2.369	4.075	3.285	4.775	1.577	3.165
<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	3.734	4.568	3.542	4.396	4.070	4.839
<i>Explain how a mathematics problem is solved</i>	10.48	7.238	8.198	7.098	12.831	6.684
<i>Engage in calendar-related activities</i>	18.27	4.725	19.03	3.624	17.61	5.434
<i>Do mathematics worksheets</i>	11.77	7.190	9.21	7.075	14.41	6.413
<i>Do mathematics problems from the textbook</i>	7.666	8.633	3.737	6.788	11.640	8.519
<i>Complete mathematics problems on the chalkboard</i>	6.902	7.066	4.666	6.187	9.291	7.211
<i>Solve mathematics problems in small groups or with a partner</i>	7.102	6.252	6.301	6.215	7.949	6.226
<i>Work on mathematics problems that reflect real-life situations</i>	8.905	6.693	7.968	6.760	9.853	6.519
<i>Work in mixed achievement groups</i>	9.638	7.718	9.978	7.872	9.407	7.521
<i>Peer tutoring</i>	6.108	6.771	5.545	6.912	6.823	6.688
<i>Do worksheets or workbook page emphasizing routing practice or drill</i>					11.60	6.729
<i>Work on problems for which there are several solutions</i>					7.200	6.570
Teacher Content Practices						
<i>Correspondence between numbers and quantity</i>	11.43	7.721	14.163	6.109	9.020	8.230
<i>Writing all numbers from 1 to 10</i>	8.933	7.876	11.296	6.892	6.882	8.236
<i>Counting by 2s, 5s, and 10s</i>	10.06	7.352	9.409	7.710	10.73	6.943
<i>Counting beyond 100</i>	6.475	7.473	6.051	7.810	7.092	7.224
<i>Writing all numbers from 1 to 100</i>	4.112	5.799	2.883	5.267	5.522	6.124
<i>Recognizing and naming geometric shapes</i>	6.480	6.316	8.326	6.898	4.994	5.439
<i>Identifying relative quantity</i>	9.701	6.695	9.880	6.799	9.606	6.679
<i>Sorting objects into subgroups according to a rule</i>	5.976	5.550	7.301	6.077	4.855	4.884
<i>Ordering objects by size or other properties</i>	5.314	5.167	6.415	5.657	4.412	4.636
<i>Making, copying, or extending patterns</i>	8.170	6.840	9.738	6.976	6.693	6.324
<i>Recognizing the value of coins and currency</i>	7.771	7.070	5.956	6.516	9.610	7.145
<i>Adding single-digit numbers</i>	11.62	7.022	8.582	6.944	14.864	5.632
<i>Subtracting single-digit</i>	10.37	7.319	6.598	6.708	14.392	5.811
<i>Place values</i>	9.192	8.313	6.647	8.631	11.733	7.118
<i>Reading two-digit numbers</i>	13.23	7.216	12.609	7.782	14.003	6.546
<i>Reading three-digit numbers</i>	6.985	7.981	5.514	7.969	8.408	7.795

<i>Mixed operations</i>	1.855	4.368	0.586	2.693	3.139	5.330
<i>Reading simple graphs</i>	7.291	6.612	7.439	6.967	7.248	6.276
<i>Performing simple data collection and graphing</i>	5.457	5.909	5.108	5.992	5.944	5.897
<i>Fractions</i>	2.656	3.927	1.761	3.391	3.622	4.268
<i>Ordinal numbers (e.g., first, second, third)</i>	7.607	7.043	8.556	7.485	6.756	6.490
<i>Using measuring instruments accurately</i>	3.351	4.292	2.717	3.905	4.088	4.679
<i>Telling time</i>	8.006	7.458	5.828	6.874	10.275	7.411
<i>Estimating quantities</i>	5.223	5.612	4.560	5.431	5.893	5.735
<i>Adding two-digit numbers</i>	3.519	5.735	0.919	3.478	6.424	6.504
<i>Carrying numbers in addition</i>	0.918	3.387	0.309	2.129	1.691	4.488
<i>Subtracting two-digit numbers</i>	2.716	5.197	0.508	2.729	5.155	6.237
<i>Estimating probability</i>	2.116	3.886	1.393	3.502	2.848	4.186
<i>Writing mathematics equations to solve word problems</i>	4.889	6.270	1.828	4.063	8.019	6.634
Class Characteristics						
<i>Class percent Asian</i>	0.038	0.079	0.121	0.055	0.102	0.048
<i>Class percent Hispanic</i>	0.130	0.236	0.168	0.114	0.140	0.100
<i>Class percent Black</i>	0.148	0.247	0.177	0.118	0.145	0.095
<i>Class percent other race</i>	0.009	0.028	0.106	0.029	0.086	0.017
<i>Class percent disabled</i>	0.082	0.102	0.153	0.072	0.124	0.058
<i>Hours per week on math</i>	3.711	1.807	3.045	1.829	4.368	1.540
<i>Hours per week in math divided achievement groups</i>	0.699	1.166	0.482	1.025	0.927	1.279
School Characteristics						
<i>Minority Percent Quintile</i>	2.620	1.530	2.720	1.551	2.690	1.549
<i>Gang problems</i>	0.294	0.456	0.333	0.471	0.275	0.446
<i>Crime problems</i>	0.380	0.485	0.422	0.494	0.354	0.478
<i>Private religious</i>	0.187	0.390	0.172	0.377	0.189	0.392
<i>Private non-religious</i>	0.025	0.157	0.042	0.201	0.024	0.153
<i>Midwest</i>	0.256	0.436	0.246	0.431	0.252	0.434
<i>South</i>	0.337	0.473	0.331	0.470	0.349	0.477
<i>West</i>	0.225	0.417	0.229	0.420	0.218	0.413
<i>Suburban</i>	0.383	0.486	0.384	0.486	0.392	0.488
<i>Rural</i>	0.231	0.422	0.209	0.407	0.229	0.420
<i>Enrollment 300 to 499</i>	0.305	0.460	0.279	0.448	0.304	0.460
<i>Enrollment 500 to 749</i>	0.273	0.446	0.291	0.454	0.252	0.434
<i>Enrollment greater than 749</i>	0.185	0.388	0.161	0.368	0.213	0.410
<i>N (Single Imputed Set)</i>	21,232		16,356		11,780	

Source: ECLS-K; Multiply imputed data

Appendix Table 2: Summary of Missing Data

	Kindergarten		1 st Grade	
	Non-missing Obs	Percent of Possible	Non-missing Obs	Percent of Possible
<i>Math score</i>	16356	100.00%	11780	100.00%
<i>Math score lag</i>	16356	100.00%	11780	100.00%
<i>Teacher Black</i>	16356	100.00%	11780	100.00%
<i>Teacher other race</i>	16356	100.00%	11780	100.00%
<i>Teacher Hispanic</i>	16356	100.00%	11780	100.00%
<i>Less than five years experience</i>	16261	99.42%	11525	97.84%
<i>More than nine years experience</i>	16261	99.42%	11525	97.84%
<i>Degree greater than BA</i>	15379	94.03%	11483	97.48%
<i>Degree MA or above</i>	15379	94.03%	11483	97.48%
<i>Certification status</i>	15715	96.08%	11402	96.79%
<i>More than two course in teaching math</i>	15307	93.59%	10925	92.74%
<i>More than two hours paid preparatory time</i>	15637	95.60%	11157	94.71%
<i>More than five hours non-paid preparatory time</i>	15908	97.26%	11525	97.84%
<i>Class percent disabled</i>	15005	91.74%	11464	97.32%
<i>Class percent Black</i>	14550	88.96%	11069	93.96%
<i>Class percent Hispanic</i>	14567	89.06%	11069	93.96%
<i>Class percent Asian</i>	14551	88.96%	11069	93.96%
<i>Class percent other race</i>	14338	87.66%	11069	93.96%
<i>Spanish exam</i>	16356	100.00%	11780	100.00%
<i>Days between exams</i>	16356	100.00%	11780	100.00%
<i>Disabled</i>	15291	93.49%	10758	91.32%
<i>Mother below high school</i>	15566	95.17%	10752	91.27%
<i>Mother high school</i>	15566	95.17%	10752	91.27%
<i>Mother some college</i>	15566	95.17%	10752	91.27%
<i>Mother BA or above</i>	15566	95.17%	10752	91.27%
<i>Income less than \$20k</i>	15803	96.62%	10289	87.34%
<i>Income more than \$100k</i>	15803	96.62%	10289	87.34%
<i>Household size</i>	15183	92.83%	10784	91.54%
<i>Number of siblings</i>	15183	92.83%	10784	91.54%
<i>Only Child</i>	15183	92.83%	10784	91.54%
<i>Welfare participant</i>	14551	88.96%	10741	91.18%
<i>Single parent household</i>	15183	92.83%	10784	91.54%
<i>Father absent</i>	15183	92.83%	10784	91.54%
<i>Mother work full time</i>	14948	91.39%	10494	89.08%
<i>Mother work part time</i>	14948	91.39%	10494	89.08%
<i>Pay tuition</i>	15098	92.31%	10745	91.21%

<i>Parental school involvement</i>	16356	100.00%	11780	100.00%
<i>How often child reads</i>	15137	92.55%	10767	91.40%
<i>Extracurricular activities</i>	16356	100.00%	11780	100.00%
<i>Children's books in home</i>	15166	92.72%	10729	91.08%
<i>Age in months</i>	16353	99.98%	11779	99.99%
<i>Female</i>	16355	99.99%	11780	100.00%
<i>Child Asian or Pacific-Islander</i>	16325	99.81%	11764	99.86%
<i>Child Black</i>	16325	99.81%	11764	99.86%
<i>Child Hispanic</i>	16325	99.81%	11764	99.86%
<i>Child other race</i>	16325	99.81%	11764	99.86%
<i>Non-English at home</i>	15772	96.43%	10940	92.87%
<i>KG repeater</i>	15291	93.49%	11168	94.80%
<i>Full day KG</i>	16356	100.00%	11780	100.00%
<i>Mother work before KG</i>	15155	92.66%	11059	93.88%
<i>Hours per week on math</i>	15772	96.43%	10835	91.98%
<i>Hours per week in math divided achievement groups</i>	14746	90.16%	8159	69.26%
<i>Count out loud</i>	16356	100.00%	11677	99.13%
<i>Work with geometric manipulatives</i>	16182	98.94%	11646	98.86%
<i>Work with counting manipulatives to learn basic operations</i>	16250	99.35%	11647	98.87%
<i>Play mathematics-related games</i>	16196	99.02%	11665	99.02%
<i>Use a calculator for math</i>	16228	99.22%	11716	99.46%
<i>Use music to understand mathematics concepts</i>	16277	99.52%	11709	99.40%
<i>Use creative movement or drama to understand mathematics concepts</i>	16236	99.27%	11692	99.25%
<i>Work with rulers, measuring cups, spoons, or other measuring instruments</i>	16260	99.41%	11660	98.98%
<i>Explain how a mathematics problem is solved</i>	16108	98.48%	11720	99.49%
<i>Engage in calendar-related activities</i>	16248	99.34%	11690	99.24%
<i>Do mathematics worksheets</i>	16219	99.16%	11735	99.62%
<i>Do mathematics problems from the textbook</i>	16144	98.70%	11687	99.21%
<i>Complete mathematics problems on the chalkboard</i>	16281	99.54%	11750	99.75%
<i>Solve mathematics problems in small groups or with a partner</i>	16262	99.43%	11748	99.73%
<i>Work on mathematics problems that reflect real-life situations</i>	16217	99.15%	11709	99.40%
<i>Work in mixed achievement groups</i>	16272	99.49%	11662	99.00%
<i>Peer tutoring</i>	16197	99.03%	11600	98.47%
<i>Do worksheets or workbook page emphasizing routing practice or drill</i>			11735	99.62%
<i>Work on problems for which there are several solutions</i>			11631	98.74%
<i>Correspondence between numbers and quantity</i>	15951	97.52%	11410	96.86%
<i>Writing all numbers from 1 to 10</i>	16120	98.56%	11603	98.50%
<i>Counting by 2s, 5s, and 10s</i>	16061	98.20%	11656	98.95%
<i>Counting beyond 100</i>	15965	97.61%	11509	97.70%
<i>Writing all numbers from 1 to 100</i>	16015	97.92%	11560	98.13%
<i>Recognizing and naming geometric shapes</i>	16144	98.70%	11655	98.94%
<i>Identifying relative quantity</i>	16193	99.00%	11702	99.34%
<i>Sorting objects into subgroups according to a rule</i>	16169	98.86%	11645	98.85%

<i>Ordering objects by size or other properties</i>	16101	98.44%	11592	98.40%
<i>Making, copying, or extending patterns</i>	16190	98.99%	11675	99.11%
<i>Recognizing the value of coins and currency</i>	16056	98.17%	11653	98.92%
<i>Adding single-digit numbers</i>	16130	98.62%	11731	99.58%
<i>Subtracting single-digit</i>	15895	97.18%	11711	99.41%
<i>Place values</i>	16011	97.89%	11629	98.72%
<i>Reading two-digit numbers</i>	16071	98.26%	11624	98.68%
<i>Reading three-digit numbers</i>	16015	97.92%	11539	97.95%
<i>Mixed operations</i>	16067	98.23%	11604	98.51%
<i>Reading simple graphs</i>	16071	98.26%	11641	98.82%
<i>Performing simple data collection and graphing</i>	16201	99.05%	11719	99.48%
<i>Fractions</i>	16076	98.29%	11619	98.63%
<i>Ordinal numbers (e.g., first, second, third)</i>	16079	98.31%	11663	99.01%
<i>Using measuring instruments accurately</i>	15990	97.76%	11524	97.83%
<i>Telling time</i>	15931	97.40%	11595	98.43%
<i>Estimating quantities</i>	15821	96.73%	11441	97.12%
<i>Adding two-digit numbers</i>	16143	98.70%	11358	96.42%
<i>Carrying numbers in addition</i>	16211	99.11%	11610	98.56%
<i>Subtracting two-digit numbers</i>	16112	98.51%	11542	97.98%
<i>Estimating probability</i>	16051	98.14%	11499	97.61%
<i>Writing mathematics equations to solve word problems</i>	16192	99.00%	11631	98.74%
<i>Sample Weight</i>	16356	100.00%	11780	100.00%
<i>Minority Percent Quintile</i>	15973	97.66%	11619	98.63%
<i>Private religious</i>	16356	100.00%	11780	100.00%
<i>Private non-religious</i>	16356	100.00%	11780	100.00%
<i>Enrollment 300 to 499</i>	16237	99.27%	11685	99.19%
<i>Enrollment 500 to 749</i>	16237	99.27%	11685	99.19%
<i>Enrollment greater than 749</i>	16237	99.27%	11685	99.19%
<i>West</i>	16356	100.00%	11780	100.00%
<i>South</i>	16356	100.00%	11780	100.00%
<i>Midwest</i>	16356	100.00%	11780	100.00%
<i>Suburban</i>	16356	100.00%	11731	99.58%
<i>Rural</i>	16356	100.00%	11731	99.58%
<i>Gang problems</i>	13917	85.41%	10507	89.19%
<i>Crime problems</i>	13960	83.70%	10548	89.54%
<i>All Variables</i>	6271	38.34%	3923	33.30%

Source: ECLS-K; Multiply imputed data

Figure 1. Tree Diagram of Possible Model/Estimation Strategies

